Supporting Information

for

Angew. Chem. Int. Ed. Z53588

© Wiley-VCH 2004

69451 Weinheim, Germany
Simultaneous Assignment and Structure Determination of Protein Backbones using NMR Dipolar Couplings

by Young-Sang Jung, Mukesh Sharma and Markus Zweckstetter

Figure S1. Overview of the Itas fold determination procedure.
**Figure S2.** Cross-validation using H$^N$-H$^N$ NOEs of the 100 lowest energy structures of ubiquitin obtained by Itas. $\Delta$ indicates the backbone root-mean-square-deviation (rmsd) between a Itas structure and the crystal structure (PDB code: 1UBQ). $\delta_{\text{NOE}}$ is the rmsd between experimental H$^N$-H$^N$ NOEs (PDB code: 1G6J) and those back-calculated from Itas structures. The linear regression coefficient is 0.79.
Appendix S3. RDC-enhanced assignment and assignment analysis.

During Itas backbone assignment is bootstrapped by best-fitting experimental RDCs to values back-calculated from Itas folds of improving quality according to

$$
\chi^2 = \sum_{l=1}^{N_{RDC}} \left( \frac{RDC(i)_l^{exp} - RDC(j)_l^{cal}}{\sigma_{lRDC}} \right)^2
$$

where $RDC(i)_l^{exp}$ is the experimental RDC of type $l$ (e.g. $^1$D$_{NH}$ or $^1$D$_{CaC'}$) of pseudo-residue $i$, $RDC(j)_l^{cal}$ is the back-calculated RDC of type $l$ of residue $j$, $N_{RDC}$ is the number of RDC types and $\sigma_{lRDC}$ is the value used for normalizing RDC deviations. $RDC(j)_l^{cal}$ is obtained by singular value decomposition. In order to take into account the low resolution of protein backbone structures especially in initial phases of Itas, $\sigma_{lRDC}$ has been optimized to $\sigma_{lRDC} = 0.28 D_a^{HN}$, where $D_a^{HN}$ is the magnitude of the alignment tensor (normalized to $^1$D$_{NH}$).

With the increase in the amount of experimental information and the improvement in tertiary structure during Itas the identification of secondary structure is continuously improved, too. Therefore, the probability of a residue belonging to a specific type of secondary structure is determined after each structure assembly step. For each residue the type of secondary structure that is present in each of the 20 lowest energy structures is extracted and the probability for a specific secondary structure is calculated from the corresponding number of structures. This secondary structure probability $p_S$ is used for a refined calculation of chemical shifts according to $\delta^{cal} = p_S \delta^{sec} + (1 - p_S) \delta^{ran}$, where $\delta^{sec}$ is the chemical shift expected when a residue is involved in an $\alpha$-helix or a $\beta$-strand (as extracted from an empirical database [2]) and $\delta^{ran}$ is the random coil value.

In each round of iteration RDC-enhanced assignment is performed using the 20 lowest energy structures generated by RosettaNmr. Assignments obtained from these 20 runs and identified as reliable [3] are evaluated for consistency. When different reliable assignments for the same position in the primary sequence are obtained for different structures, the corresponding assignments are discarded. When a pseudo-residue is assigned reliably for a specific structure, it is, however, also assigned ‘reliably’ to a different residue for a different structure, the assignment is removed. When the overall number of reliable assignments does not improve from one to the next iteration step, the criteria for identifying assignments are relaxed and ASS$^{\text{global}}$ assignments [3] are taken instead of reliable assignments for the 20 top scoring structures. This is necessary as slight improvements in Itas folds sometimes do not lead to an increase in the number of reliable assignments and Itas would be trapped. As there will be significant variation in ASS$^{\text{global}}$ assignments across the 20 top scoring structures, ASS$^{\text{global}}$ assignments that were obtained for most of the 20 top scoring structures are selected. From these assignments those are further removed for which a pseudo-residue is assigned to more than one residue.